DOCUMENT RESUME

ED 315 438 TM 014 461

AUTHOR Cameron, Robert G.
TITLE The Common Yardstick: A Case for the SAT.
INSTITUTION College Entrance Examination Board, New York, N.Y.
PUB DATE 89
NOTE 30p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS College Bound Students; *College Entrance Examinations; *Educational Assessment; Higher Education; High Schools; Standardized Tests; Test Reliability; *Test Use; Test Validity
IDENTIFIERS *Scholastic Aptitude Test

ABSTRACT

The contribution of the Scholastic Aptitude Test (SAT) to education is described, and the most frequently asked questions about its use are addressed. The SAT became a prominent tool in college admissions when the members of the College Board recognized in the 1930's that the Board's program of very specific subject-matter tests was no longer appropriate or welcome since educational reforms had freed the high school curriculum of rigid academic requirements. Once the limitations of subject-specific examinations were recognized, the virtues of a test of general reasoning and thinking skills, such as the SAT, also became obvious. Benefits of the SAT covered in this discussion include those related to uses by parents and students, benefits to high schools and colleges, state-level benefits, and benefits to society at large. Particular measurement characteristics of the SAT reduce sex and ethnic biases. Finally, the effects of test coaching on SAT design and the measures of the validity and effectiveness of each SAT item are addressed. (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED315438

The Common Yardstick

A Case for the SAT

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Robert G. Cameron

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM014461

The College Board

The Common Yardstick

A Case for the SAT®

Robert G. Cameron
Executive Director
Research and Development
The College Board

The Common Yardstick: A Case for the SAT

ERRATUM

p. 13 Table 6...Freshman GPA, Other Students should read 2.60, not 2.06

College Entrance Examination Board, New York, 1989

The College Board is a nonprofit membership organization committed to maintaining academic standards and broadening access to higher education. Its more than 2,600 members include colleges and universities, secondary schools, university and school systems, and education associations and agencies. Representatives of the members elect the Board of Trustees and serve on committees and councils that advise the College Board on the guidance and placement, testing and assessment, and financial aid services it provides to students and educational institutions.

Copyright © 1989 by College Entrance Examination Board. All rights reserved.
The College Board, Scholastic Aptitude Test, SAT, and the acorn logo are registered trademarks of the College Entrance Examination Board.
Printed in the United States of America.

Contents

Foreword *v*

Introduction *1*

"Uncommon" Schooling *1*

The Need for an Objective Measure *2*

The Benefits of the SAT *3*

How Students and Parents Use the SAT *3*

Benefits to High Schools *4*

Benefits to Colleges *4*

Benefits to States *7*

Benefits to Society *7*

Beneficial Measurement Characteristics *9*

The SAT and Ethnic Minorities *9*

The SAT and Gender *15*

The SAT and Coaching *18*

Choice of SAT Questions *19*

Summing Up *21*

Glossary *23*

Bibliography *25*

Foreword

The Scholastic Aptitude Test (SAT) is probably known to more Americans than any other examination except the one they take to qualify for a driver's license. The SAT is taken each year by two-fifths of our high school graduates as a rite of passage from secondary school to higher education. It is often the most visible part of the college admissions process—yet its content, use, and role in that process are often misunderstood. This publication helps to describe the SAT's contributions to education and to answer the most frequently asked questions about its use.

The author, Robert G. Cameron, is executive director of Research and Development at the College Board. The text is a distillation of his many years of experience in counseling and admissions, his knowledge of measurement research, and hundreds of conversations with those who take or use the SAT. I believe that you will find *The Common Yardstick: A Case for the SAT* an interesting and revealing account of why the SAT is so important in our unique educational system—and why it is needed as much today as ever before.

Donald M. Stewart
President
The College Board

Introduction

The Scholastic Aptitude Test (SAT) was first taken by college-bound students in 1926. Since that time, the test has gradually evolved into its current form, one that has remained relatively the same since 1975, when the Test of Standard Written English (TSWE) was introduced as a component. The SAT is among the most researched of tests. This publication, written for the nonexpert, cites the results of some of this research. (A glossary of technical terms is provided.) The purpose is to demonstrate the value of the SAT to students and their parents, to high schools and systems of schools, and to colleges, states, and society. It answers often-asked questions such as:

- What does the test measure?
- Why does it measure what it does?
- What is its role and purpose in college admissions?

"Uncommon" Schooling

When the founders of this nation delegated to the states the power to regulate education, they created a decentralized system that brought to the fore local decision making regarding the content and the emphasis of the curriculum. This tradition has remained with us. Decisions concerning what will be taught, what will be emphasized, and what will ultimately be measured are largely made by teachers. State and local control of content and standards creates graduates who have had a variety of experiences and whose grades may vary widely because of very different grading standards. A teacher's determination of whether students have met specific standards is sometimes influenced by subjective judgment. The SAT provides a corrective influence on grades and standards, as would any standardized test that is commonly used in college admissions. That influence, however, is indirect, as was observed during the 1970s when high school grades became increasingly inflated while SAT averages were falling. Public concern over the decline in SAT scores led, in part, to the current school reform movements and somewhat more rigorous grading standards in high schools.

The SAT fits well into the nation's curricular heterogeneity in that it presumes no particular pattern of course work or course content. The abilities it measures are generalized abilities needed for performance across wide-ranging undergraduate curriculums. The curricular paths followed in elementary and secondary school in preparation for the SAT can be diverse, provided students study academic courses and include college preparatory mathematics among their studies.

The Need for an Objective Measure

When the Constitution made schooling the concern of the states, it created a situation in which curriculum and standards could differ from one state and one school district to another. The need arose for a common yardstick to measure school outcomes in college admissions, and the SAT has become that yardstick. The test is fair to diverse populations of high school graduates because the generalized abilities tested do not presume rigid patterns of course preparation.

The SAT became a prominent tool in college admissions when the members of the College Board recognized in the 1930s that the Board's program of very specific subject-matter tests was no longer appropriate or welcome since the progressive and other educational reforms freed the high school curriculum of rigid academic requirements. The early program had an adverse impact on the high school curriculum in that teachers taught to the tests. The curriculum-neutral SAT had the advantage of greater accuracy of measurement than the written subject-matter examinations it gradually replaced. Today's, and tomorrow's, post-technological society demands citizens who can read with depth and comprehension, who can reason and discern complex relationships, and who can think critically. The SAT measures these skills.

Other nations have come to recognize the limitations of specific subject-matter examinations and have looked to the United States for a model test that measures generalized reasoning and thinking skills. They have found that model in the SAT, and through the International Association for Educational Assessment they have developed the International Test of Developed Abilities, which measures mathematical and verbal abilities in four languages. Each language version is equated to the SAT. Great Britain, in an effort to increase access to its higher education institutions, is considering eliminating the A-Level and O-Level examinations in favor of tests of generalized abilities.

The Benefits of the SAT

How Students and Parents Use the SAT

Students and parents use the SAT or its preliminary version, the Preliminary Scholastic Aptitude Test/National Merit Scholarship Qualifying Test (PSAT/NMSQT), for a variety of purposes in facilitating the transition from high school to college. First, the SAT is used to gauge individual abilities vis-à-vis the competition in college. Parents and students often do not know how rigorous the grading standards of their local school are. SAT scores are a gauge by which students can determine their relative standing in the academic competition for admission to a given college. The scores can help a student focus on a manageable number of colleges to which admission seems desirable and probable, given the student's abilities.

Students who take the SAT and give permission to have their names released to higher education institutions receive information from colleges that are interested in them. This exchange provides a means by which students can broaden their choices and consider educational opportunities about which they might otherwise be unaware.

The SAT provides information about students' strengths in verbal and mathematical areas, and in composition through the Test of Standard Written English. Students can begin to consider their college major and career choices in a context of comparison against others who choose those majors or careers.

The SAT becomes an "advocate" for those students coming from high schools with very rigorous grading standards and whose grade-point average may indicate lower college potential than the objective measure of the SAT. Conversely, the SAT can identify students who are likely to encounter academic difficulties; the scores can indicate a need for developmental or remedial studies in some areas. Students in remote or relatively unknown high schools can make their promise known to colleges anywhere in the country.

Benefits to High Schools

The SAT is a useful tool for teachers and counselors giving guidance to high school students. It can be used to assist students in choosing college majors and thinking about careers. It can help students in selecting colleges where they are likely to find the academic competition congenial and in scaling up or down students' and parents' unrealistic expectations of the colleges to which students plan to apply. The PSAT/NMSQT and the SAT can help identify talent to be encouraged and nurtured. Frequently students disguise their talents by nonconformance in the classroom or by poor work habits. The PSAT/NMSQT, particularly if taken in the sophomore year, can identify talented students who otherwise might not receive the kind of encouragement they need to proceed to higher education. Students capable of studying college-level courses while in high school can be identified and guided to advanced courses.

The PSAT/NMSQT, which offers schools a Summary of Answers for each student, can help schools evaluate their curriculum by identifying possible gaps or weaknesses, and by assessing the achievements of students by a single objective measure. The test results can also be used to report the accomplishments of students to the public, to colleges (in the form of high school profiles), and to state education authorities. Although not all students take the SAT, in many high schools most college-bound students do. Aggregate results can accurately describe the scholastic abilities and achievements of those students.

Benefits to Colleges

The SAT and the Student Descriptive Questionnaire (SDQ) that accompanies it provide colleges that receive score reports from students with information about the verbal and mathematical abilities, the academic and extracurricular backgrounds, and the educational and career aspirations of prospective applicants. Students who give permission to be included in the College Board's Student Search Service (SSS) constitute a pool of college aspirants in excess of 2,330,000 (1,455,000 in the PSAT/NMSQT pool and 875,000 in the SAT pool). Thus, students taking the SAT and PSAT/NMSQT create two pools of potential applicants, the smaller one consisting of those interested enough in a college to send their score report, and the larger consisting of students who may be persuaded to apply. These recruitment pools help colleges achieve their affirmative action and enrollment goals, and enable them to target their efforts at qualified students.

The SAT is the common yardstick with which colleges can compare the abilities of their applicants. The high school record alone does not enable

this because of the variation in grading standards from high school to high school and from teacher to teacher. The SAT in combination with the high school record (HSR) improves the accuracy of prediction of college grades. The median correlation of high school record and freshman grade-point average (GPA) for a sample of 685 colleges is .48, while the SAT correlation is .42. The SAT in combination with HSR correlates .55—a 15 percent improvement over high school record alone. The improvement would be greater, except that students self-select colleges based in part on their SAT scores. Students "use up" some of the power of the SAT when they consider their scores in selecting colleges appropriate to their academic potential. Students who are at the extremes of the ranges of enrolled students are more likely not to apply. This self-selection usually results in restriction of the range of SAT scores at any given college, thus lowering the observed correlation of the SAT with college GPA*. In other words, if colleges were to receive applications from students whose range of test scores mirrored that of the national test-taking population, they would observe a higher correlation between the SAT and college grades.

Table 1 displays validity data from a group of 21 colleges at which the range of SAT scores and high school record was not restricted. The average standard deviation of SAT-verbal plus SAT-mathematical was the same as the total population taking the test. In other words, the average SAT-verbal and SAT-mathematical scores and HSR were approximately the same as those of the test-taking population. At those colleges with no restriction of the range, the SAT-sum and the multiple correlation of the SAT with freshman GPA was higher than for HSR alone (see footnote, Table 1). The increment in correlation was .10, or an 18 percent improvement. Restriction of the range comes about by student self-selection of colleges and by college selection from the students who have selected them. Consequently, the correlation of the SAT with freshman GPA is reduced by student and institutional selection.

After admission to college, the SAT, TSWE, and the student responses to questions about high school courses taken and grades earned help in counseling students about majors and programs of study, and in placing students in appropriate levels of courses, from remedial to honors. Sensitive and timely counseling and placement activities can contribute positively to student progress and satisfaction during the first, critical months of college. Students with a college GPA below 2.0 during freshman year have a high

*For example, colleges with severer restriction of the range (standard deviations below 80) observe an average SAT correlation with college grades of .36, whereas those colleges with standard deviations of 100 or higher observe average correlations of .49.

Table 1: Average Validity Data for Colleges with No Observable Restriction of Range

Number of Colleges	21
SAT Means	
SAT-V	459
SAT-M	477
SAT-Sum*	936
SAT Standard Deviations	
SAT-V	111
SAT-M	114
SAT-Sum*	208
High School Record Means	
High School Rank (20–80 scale)	57.00
High School GPA (0–4 scale)	3.02
Mean Freshman GPA	2.64
Proportional Contributions	
SAT-V	31%
SAT-M	20%
HSR	48%
Correlations with Freshman GPA	
SAT-V	.53
SAT-M	.50
SAT-Multiple**	.57
HSR	.55
SAT-V, SAT-M, HSR Multiple	.65
SAT Increment	.10

Adapted from Donlon 1984, p. 147.

* The SAT sum is a simple sum of SAT-verbal (SAT-V) and SAT-mathematical (SAT-M).

** The SAT-multiple correlation with freshman GPA is based on the optimal weighting of the two scores.

probability of dropping out of college. Placement in the appropriate courses or programs can help students meet their academic goals and increase the college's retention rate.

Academic standards are set and maintained by college faculty. Objective indices of the standards are, for example, acceptance rates in graduate and professional programs, accreditation, and standardized test scores. The SAT can help colleges set realistic academic and admissions standards by describing the academic abilities of their applicants on a common scale. Admissions decisions can be guided in part by this readily understood and objective standard. Applicants' selection of colleges is guided by the test score information provided by those institutions and by the way in which the applicants perceive they meet each college's "standard." Although meeting the test score standard should be only one of many considerations in selecting a college, it is doubtful that the "great sorting" of 1.5 million students a year into higher education could take place in the orderly fashion it does without a standardized measure such as the SAT.

Benefits to States

States benefit from the SAT in a variety of ways. States' summary reports of test scores and academic and background information enable state education officials, legislators, and the public to track the abilities, accomplishments, and aspirations of the college bound over time. Although the College Board cautions against state-by-state score comparisons because the percentage of high school graduates taking the SAT varies widely by state (from 3 percent to 69 percent), trends over time within a state can reveal how its students are progressing toward educational goals. Among the trends that might be noted are broadened participation by minorities in core academic subjects, increased study in honors courses, and shifts in demand for certain college majors or graduate degrees.

Benefits to Society

The SAT and SDQ benefit society by enabling a compilation of trends and by making possible an annual assessment about the majority of the nation's college bound. The first alarm that America's schools were slipping sounded when the College Board announced in 1975 a decade-long decline in SAT scores. The reaction—great public concern, debate, and later,

reform—demonstrated that the common yardstick could serve a new purpose.

Admissions tests such as the SAT serve the public interest because they are sponsored by voluntary associations *at no cost to the taxpayer*. Society and individuals benefit from the information they generate, and the great sorting takes place without governmental intervention or expense. In the best academic tradition, college faculties are free to shape policy and admissions criteria without governmental interference. Each year over 1.5 million students distribute themselves into over 3,000 postsecondary institutions. Literally millions of individual decisions, the majority made outside the admissions office, support this process. For most of these students, test scores play a prominent role in their decision making.

Beneficial Measurement Characteristics

We have seen how the SAT benefits individuals, institutions, and society. Its measurement characteristics make it an invaluable tool in education.

The SAT measures generalized verbal and mathematical reasoning abilities that are applicable across the undergraduate curriculum. If the SAT were irrelevant or too narrow a gauge of the tasks students need to perform in college, one would expect to find low average validity for it in some, if not all, areas of college study. Table 2 contains average validity data by area of college study. The proportional contribution of SAT-verbal and SAT-mathematical scores varies from program to program. SAT-verbal contributes most significantly to liberal arts; SAT-mathematical, to engineering. The contribution of both scores to the prediction of grades ranges from 42 percent to 46 percent. This consistency of contribution across varying undergraduate programs suggests that the SAT measures a wide range of generalized abilities relevant to predicting performance effectively across academic programs.

The SAT and Ethnic Minorities

Some people believe that differences in mean scores between groups are evidence of test bias. Those people fail to examine the question of whether differences in backgrounds of the groups might explain the differences. Table 3 displays test score averages and Table 4 displays academic and socioeconomic background for the major ethnic subgroups taking the SAT. Socioeconomic factors are included because educational achievement is positively correlated with parents' education and family income; academic background is included because prior educational achievement forecasts future educational achievement.

Table 2: Average Validity Data by Area of Study

	Business	*Liberal Arts*	*Engineering*	*Science*	*Education*	*Nursing*
Number of Colleges	100	96	77	64	61	27
SAT Means						
SAT-V	460	482	505	500	446	468
SAT-M	515	502	600	558	469	489
Proportional Contributions						
SAT-V	21%	27%	14%	15%	25%	23%
SAT-M	25%	15%	33%	31%	22%	24%
HSR	54%	58%	53%	54%	54%	54%
Correlations with Freshman GPA						
SAT-V	.30	.35	.25	.32	.36	.37
SAT-M	.32	.29	.34	.36	.34	.39
SAT-Multiple	.37	.39	.37	.41	.42	.46
HSR	.43	.48	.42	.46	.49	.50
SAT-V, SAT-M, HSR Multiple	.52	.55	.51	.55	.57	.60

Adapted from Donlon 1984, p. 155.

Table 3: SAT Averages by Ethnic Group, 1985

	SAT-V	*SAT-M*
Asian American	404	518
Black	346	376
Mexican American	382	426
Puerto Rican	368	409
White	449	491
All Students	431	475

Source: Ramist and Arbeiter 1986.

The SAT-verbal averages for each ethnic minority group are below those of whites. The same is true for SAT-mathematical averages for all minority groups except Asian Americans, whose average is 27 points above that of whites.

An examination of academic background offers some explanation for the differences. Smaller percentages of minorities in high school are enrolled in college preparatory programs, and with the exception of Asian Americans, minorities study fewer units or years of academic subjects. The additional study by Asian Americans tends to be concentrated in math and science, which helps to explain their higher SAT-mathematics scores. Thus, the nature of the high school program and the number of academic subjects studied has a positive relationship to test score averages.

The positive relationship between socioeconomic status, educational attainment, and test scores has been well documented in the research literature. The differences shown in Table 4 in level of parents' education and median income among ethnic groups taking the SAT are consistent with the findings of earlier studies.

There is positive correlation between income level and standardized test scores. Families with above-average incomes are more likely to be headed by better educated parents. These parents are more likely to have college aspirations for their children and are more likely to encourage their children in school. Their higher incomes make it possible to choose to live in communities with good schools. These families are likely to have greater discretionary income to expend on educational and cultural activities that enhance

Table 4: SAT Takers, 1985: Academic and Socioeconomic Background

	College Preparatory Program	*Nonpublic School*	*Years of Academic Subjects**	*Father's Education (total years)*	*Mother's Education (total years)*	*Median Income ($)*
Asian American	76%	20%	17.18	14.5	13.6	26,400
Black	65%	13%	15.41	12.4	12.8	17,100
Mexican American	69%	20%	15.47	12.1	11.9	20,500
Puerto Rican	64%	23%	16.01	12.1	12.0	17,000
White	81%	20%	16.70	14.3	13.6	34,700
All Students	79%	20%	16.55	14.1	13.6	32,200

Source: Ramist and Arbeiter 1986.

* Biological and physical sciences, English, foreign language, mathematics, history, and social studies.

educational achievement. As Table 5 demonstrates, higher SAT scores are related to higher high school grade-point average and class rank.

Does the SAT underpredict college performance for minorities? A study at the University of California, Riverside, using SAT scores and high school records, found that Asian Americans' actual college grade-point averages were as close to predicted GPAs as were those for whites. Table 6 shows that black students at predominantly white colleges achieved actual GPAs below those predicted using the total group equation. In other words, the SAT and high school record overpredicted college performance for blacks. However, over- and underpredictions for important subgroups can be avoided by studying these groups separately and by using a prediction equation unique to each group.

Table 6 also displays how the SAT contributes proportionately to the prediction of college grades for blacks and whites. At predominantly white colleges the SAT contributes 46 percent to the prediction for white students whereas the contribution on the SAT for blacks is 58 percent. At predominantly black colleges the SAT contributes 56 percent of the prediction for blacks. In other words, the SAT more efficiently predicts college grades for blacks at both types of colleges than does high school record.

Table 5: Mean SAT Scores by High School Grade-Point Average and Class Rank

	SAT-V	*SAT-M*
Rank		
Top Tenth	518	585
Second Tenth	456	511
Second Fifth	418	461
Third Fifth	380	409
Fourth Fifth	358	380
Fifth Fifth	353	374
Grade-Point Average		
A+ (97–100)	559	629
A (93–96)	512	576
A– (90–92)	482	541
B (80–89)	420	462
C (70–79)	365	395
D, E, or F (below 70)	350	381

Source: College Board, *College-Bound Seniors, National Report, 1987.*

Table 6: Average Validity Data for Black Students at Predominantly White and Predominantly Black Colleges

	Predominantly White Colleges		
	Black Students	*Other Students*	*Predominantly Black Colleges*
Number of Colleges	11	11	11
SAT Means			
SAT-V	390	463	314
SAT-M	416	509	343
Freshman GPA			
Actual Mean	2.03	2.06	2.19
Predicted Mean Using Total Group Equation	2.28	—	—
Difference	−0.25	—	—
Proportional Contributions			
SAT-V	34%	22%	25%
SAT-M	24%	24%	31%
HSR	42%	54%	44%
Correlations with Freshman GPA			
SAT-V	.24	.33	.31
SAT-M	.23	.34	.32
SAT-Multiple	.30	.39	.38
HSR	.25	.42	.36
SAT-V, SAT-M, HSR Multiple	.37	.49	.47
SAT Increment	.12	.07	.11

Adapted from Donlon 1984, p. 158.

Bias in test use is perhaps the most controversial issue concerning the use of test scores for admission of minorities to higher education. Enrollment in higher education and test-taking has increased steadily since 1980 for all minorities except blacks. Overall, the acceptance rates for minority applicants to four-year colleges were about the same as acceptance rates for all applicants. The acceptance rates of minority applicants at the "most selective" public and private four-year colleges were *higher* than rates for all applicants (Breland, Wilder, and Robertson 1986). Thus, it does not appear that tests are being used to reject a larger proportion of minority applicants, or that minority students do not take tests required for admission.

Does the SAT, by virtue of its test questions or components, unfairly represent any particular subgroup of the test-taking population? Is the measurement biased against blacks, for example?

Measurement bias is described as biased selection of test questions or construction of the test components that results in an unfair representation of the competencies measured for particular subgroups of the test-taking population. Some critics have indicted the SAT as reflective of white middle-class vocabulary, values, and competencies that are beyond the experience of ethnic minorities and, furthermore, are irrelevant to college instruction. If this were the case, one could expect the SAT to contribute little to prediction in predominantly black educational environments.

Table 7 displays data for predominantly black colleges and all colleges. If there were an unfair representation of the competencies measured for blacks, one would expect to find the SAT a less efficient predictor of college GPA at predominantly black colleges because of a lack of fit with the competencies demanded by the instruction and curriculum. Yet, the proportional contribution of the SAT to prediction for students at predominantly black colleges is greater than that of the high school record. And, the SAT more strongly predicts for black students than does the high school record at predominantly white colleges. Measurement bias does not appear to be evident for black students at predominantly black institutions or at predominantly white institutions (Tables 6 and 7).

Table 7: Average Validity Data for Predominantly Black Colleges and All Colleges, 1964–81

	Predominantly Black Colleges	*All Colleges*
Number of Colleges	22	685
SAT Means		
SAT-V	327	455
SAT-M	351	487
Proportional Contributions		
SAT-V	28%	26%
SAT-M	24%	20%
HSR	48%	54%
Correlations with Freshman GPA		
SAT-V	.31	.36
SAT-M	.30	.35
SAT-Multiple	.36	.42
HSR	.37	.48
SAT-V, SAT-M, HSR Multiple	.48	.55
SAT Increment	.11	.07

Adapted from Donlon 1984, p. 157.

The validity data in the preceding tables suggest that the SAT is free from structural and content bias. The test contributes significantly to prediction at predominantly white and predominantly black colleges and for black students at both types of institutions, and it is useful in predicting performance across undergraduate academic programs.

The SAT and Gender

Differences in average scores between men and women have been observed for many years. Women have scored lower on the mathematical section by about 50 points for the last several decades. In contrast, women's verbal scores were higher than men's until 1973, when women began to score lower. The gap between the verbal scores continued to widen slowly through the 1980s and was 13 points in 1988; the gap in mathematical scores has narrowed slightly to 43 points.

In 1973, for the first time the majority (52 percent) of the test-takers were females. Thus, some of the differences in female and male averages can be attributed to a larger female population (since whenever a test-taking pool grows, average scores for that pool decline).

The data in Table 8 suggest that the higher percentage of males enrolled in college preparatory programs and the fact that males study slightly more academic courses in high school contributes to the difference as well. As we have seen previously, socioeconomic factors have been correlated with educational attainment. Males taking the SAT come from families where the educational level of both parents is slightly higher than that of women test-takers. Women are more likely to come from a family where neither parent attended college. Family income is found to be higher for men as well. Differences in population size, academic background, and socioeconomic status help to explain the difference in mean scores. A recent study (Burton, Lewis, and Robertson 1988) found that when adjusted for ethnic group, socioeconomic status, high school curriculum, and intended college major, the difference in SAT-verbal average scores between females and males was reduced from 12 points to 2 points. The difference in the SAT-mathematical average was reduced from 53 points to 30 points.

Table 9 presents average validity data by sex. The major conclusions to be drawn from these data are: (1) the SAT contributes more to the prediction of college grades of females; and (2) the combination of high school record and SAT predicts more accurately for females than for males at the 574 colleges participating in the studies.

Test critics have charged that the SAT underpredicts college GPA for women, that, in other words, women's actual college GPA is higher than the

Table 8: Academic Background by Sex for SAT Takers, 1985

	College Preparatory Program	*Nonpublic School*	*Years of Academic Subjects**	*Father's Education Level*	*Mother's Education Level*	*Median Income ($)*
Female	78%	18%	16.46	14.0	13.5	31,200
Male	80%	21%	16.67	14.3	13.6	33,200

Source: Ramist and Arbeiter 1986

* Biological and physical sciences, English, foreign language, mathematics, history, and social studies.

Table 9: Average Validity Data by Sex, 1964–81

	Females	*Males*	*Female–Male Difference*
Number of Colleges	574	511	
SAT Means			
SAT-V	473	467	+6
SAT-M	485	520	−35
Proportional Contributions			
SAT-V	27%	25%	+2%
SAT-M	22%	22%	0%
HSR	51%	54%	−3%
Correlations with Freshman GPA			
SAT-V	.39	.32	+.07
SAT-M	.39	.33	+.06
SAT-Multiple	.46	.38	+.08
HSR	.49	.44	+.05
SAT-V, SAT-M, HSR Multiple	.57	.51	+.06
SAT Increment	.08	.07	+.01

Adapted from Donlon 1984, p. 164.

predicted average. This charge is based upon results of a College Board research study that are shown in Table 10. For the total freshman class in 41 colleges, actual grades for women do tend to be slightly higher than their predicted grades. But the difference between actual and predicted grades is .04 of a grade point, and this small difference can be eliminated by using separate prediction equations for each sex, rather than a single equation based on the total group. The *Guidelines on the Uses of College Board Test Scores and Related Data* urges that colleges "take into appropriate consideration prediction of performance for applicant subgroups—men and women, ethnic groups, international students, adults, and those interested in different academic programs—in developing equitable admissions policies and practices."

Table 10: Predicted versus Actual College GPA By Sex and by Discipline, 1980

	Men			*Women*		
	N	*Predicted Grades**	*Actual Grades*	*N*	*Predicted Grades**	*Actual Grades*
Engineering	1,145	2.55	2.54	323	2.53	2.58
Pharmacy	44	2.48	2.66	67	2.57	2.45
Health Science	46	2.89	2.93	91	2.79	2.77
Biology	208	2.82	2.85	230	2.70	2.68
Science Majors	223	2.69	2.69	73	2.61	2.61
Science and Math	53	2.93	2.83	112	2.93	2.98
Agriculture and Home Economics	79	2.52	2.50	140	2.59	2.60
Business	854	2.73	2.71	779	2.78	2.81
Liberal Arts	406	2.53	2.50	510	2.63	2.65
Humanities	62	2.88	2.80	114	2.86	2.90
Speech and Drama	22	2.81	2.65	49	2.85	2.93
Communications	92	2.74	2.68	68	2.76	2.84
Fine Arts	45	3.04	2.97	62	3.05	3.10
Average (41 Freshman Classes)	817	2.57	2.52	845	2.64	2.68

Adapted from Clark and Grandy 1984.

* Based on three-variable prediction equation for *Total Group* (SAT-V, SAT-M, and high school average or rank).

The SAT and Coaching

In the best of possible worlds, college admissions tests would be impervious to short-term preparation or coaching. In this world, students would not be concerned with preparing for the "college boards" except tangentially—as they pursued their high school studies and outside reading. Because admissions and other standardized tests measure what has been learned, they are not impervious to the effects of instruction. The demand for special preparation for admissions tests, whether conducted as a part of the school curriculum, as an extracurricular activity, or as a commercial venture, will be proportional to the perceived importance of the test results. Demand will be independent of what the test purports to measure and independent of the effectiveness of the preparation.

How coachable is the SAT? What advantages do the coached gain? Surveys of the literature and meta-analysis of studies of coaching for educational tests reveal that the SAT is the test least subject to the effects of short-term instruction or coaching. The effects were small (DerSimonian and Laird 1983; Kulik, Bangert-Drowns, and Kulik 1984).

The College Board concludes that so far no one has demonstrated that coaching is sufficiently effective for all students to recommend spending significant amounts of time or money on it. Students can accomplish at least as much in school or on their own. Students appear to have the same opinion. A recent survey of students found that they averaged 10 hours or less and spent no money in preparing to take the SAT (Powers 1988).

All students should be familiar with the test format, item types, knowledge requirements, test-taking strategy, and so on before taking the SAT. This information is contained in *Taking the SAT*, which is provided free of charge to all students who register for the test.

Beyond that, well-verified research shows that not all coaching works (certainly not for all students) and that, on the average, the effects are relatively modest. Some students who participate in coaching activities do not improve their scores. Others who do nothing but take the SAT a second time enjoy score gains comparable to the unverified claims of commercial coaching companies.

The coaching that is most effective appears to involve activities very much like rigorous course work. These same activities can be pursued by students as part of their school curriculum or as extracurricular activities. Therefore, it should not be necessary to spend a lot of money on special courses to improve scores.

A suggested alternative to the "coachable" SAT is an achievement test. It is argued that achievement tests are fairer to the poor and ethnic minorities. There is no evidence available to support or refute this assertion. However, substitution of achievement tests for the SAT would probably have the

opposite effect from that intended. Research on the effects of coaching or short-term instruction by type of test shows that achievement tests are more responsive to coaching than are aptitude tests. If the advantaged can afford to pay for coaching and the coaching is for achievement tests, the outcome is likely to be a widening of the gap in test score averages between the poor and ethnic minorities and the advantaged.

Choice of SAT Questions

If a test is to achieve and retain its validity and effectiveness, each question must play its expected role in contributing to the total score. Under actual test conditions, each question must perform as intended. Each question must prove to be as difficult or easy as the test specifications demand of it; otherwise, the meaning of a given score will not remain comparable from one edition of the test to the next or over time.

SAT questions are "pretested" by test-takers under actual test-taking conditions. This procedure helps to ensure that students are making their best effort to answer the questions and that the "pretest" statistics will not calibrate a question as more difficult than it is when used later in an operational edition.

These "pretest" statistics must be compiled from large and diverse populations in order to assure that they are representative of diverse test-takers who will have their test scores based on these questions. In order to construct tests that are as free from any potential bias as possible, a sufficient number of responses is needed from the various racial and ethnic groups; these numbers are found in large pretest populations. Each test question is analyzed to determine if a question functions in the same way for minority students compared to majority students, and for men compared to women. Several editions of the SAT are made public each year, thereby increasing the number of new questions to be written, screened by experts for potential unfairness, and "pretested" and included in operational editions by the committee overseeing the test.

The screening, review, "pretest," and statistical analysis process helps to ensure that the individual test questions and the test as a whole are as fair and accurate as possible.

Summing Up

Over 1,600 colleges and universities use the SAT for a variety of purposes. Its broadened use (over 435 colleges have adopted it in the last 10 years) is evidence that admissions professionals find it a very useful tool in admissions, recruitment, and retention. The validity of the test in forecasting college grades across the undergraduate curriculum has been demonstrated. It contributes more to the prediction of college grades for blacks than the high school record. Women's grades are predicted more accurately by the SAT than are men's. In short, its validity has been established for diverse undergraduate programs and for various subgroups.

The generalized thinking and reasoning abilities measured by the test are appropriate measures of knowledge gained through the welter of high school curriculums in this nation. Its questions are relatively curriculum-free, with minor influence from course content and instruction.

Two-thirds of the 1.5 million students who enter college each year do so with the help of the SAT. This "great sorting" is accomplished without governmental intervention or expense. These students contribute to the annual record of the educational achievements, personal background, and aspirations of our college-bound youth that helps to guide public policy and our nation's decisions about secondary education.

Glossary

Correlation. The extent to which two scores or measures tend to vary concomitantly; the degree of relationship between two variables such as test scores and college grade-point average. The customary index for expressing the degree of relationship is the coefficient which can range from −1.00, showing complete negative relationship, through zero, indicating no relationship, to +1.00, showing complete positive relationship.

Median correlation. See median.

Observed correlation. The correlation obtained in an analysis of relationships.

Mean. Arithmetic average; the sum of a set of scores (or other values) divided by the number of scores (or values).

Median. The middle score in a distribution; the fiftieth percentile; the point that divides the group into two equal parts (half of the group of scores falls below the median and half above it).

Meta-analysis. The conversion of individual study results to a common metric such as a fraction of a test's standard deviation (effect size). Statistical analysis of effect sizes can help to avoid erroneous conclusions when study sample sizes are small and the number of studies large. Estimates of the magnitude of the relationship found in the collection of studies are provided.

Prediction equation. An equation for calculating a predicted college grade-point average for an individual student. The equation is the product of a statistical analysis called a validity study of the relationships of high school academic record and admissions test scores with college grades.

Standard deviation. A measure of the variability of a distribution about its mean (average). In distributions of test scores, for example, a small standard deviation would indicate a tendency of scores to cluster about the mean; a large standard deviation would indicate a wide variation in scores. In a normal distribution, approximately 68 percent of the cases lie within a distance

measured by one standard deviation on either side of the mean, and approximately 95 percent of the cases within a distance measured by two standard deviations on either side of the mean.

Validity. The extent to which a test measures what it is supposed to measure. There are different kinds of validity depending on the type of test and its purposes, and different kinds of evidence are appropriate for appraising different kinds of validity. Predictive validity—how the test predicts college grades—is commonly associated with admissions tests.

Bibliography

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. 1985. *Standards for Educational and Psychological Testing*. Washington, D.C.: American Psychological Association, Inc.

Breland, H. M. 1979. *Population Validity and College Entrance Measures*. Research Monograph Number 8. New York: College Entrance Examination Board.

Breland, H. M., G. Wilder, and N. J. Robertson. 1988. *Demographics, Standards and Equity: Challenges in College Admissions. Report of a Survey of Undergraduate Admissions Policies, Practices and Procedures*. New York: American Association of Collegiate Registrars and Admissions Officers, The American College Testing Program, The College Board, Educational Testing Service, and the National Association of College Admissions Counselors.

Burton, N., C. Lewis, and N. Robertson. 1988. *Sex Differences in SAT Scores*. College Board Report 88-9. New York: College Entrance Examination Board.

Cameron, R. G. In press. Issues in Testing Bias. *College and University*.

Clark, M. J., and J. Grandy. 1984. *Sex Differences in the Academic Performance of Scholastic Aptitude Test Takers*. College Board Report 84-8. New York: College Entrance Examination Board.

College Board. 1986. *Measures in the College Admissions Process*. New York: College Entrance Examination Board.

College Board. 1987. *College-Bound Seniors: National 1987 Profile of SAT and Achievement Test Takers*. New York: College Entrance Examination Board.

College Board. 1988. *ATP Guide for High Schools and Colleges*. New York: College Entrance Examination Board.

College Board. 1988. *Guidelines on the Uses of College Board Test Scores and Related Data*. New York: College Entrance Examination Board.

DerSimonian, R., and N. Laird. 1983. Evaluating the Effect of Coaching on SAT Scores. *Harvard Education Review* 53:1–15.

Donlon, T. F. 1984. *The College Board Technical Handbook for the Scholastic Aptitude Test and Achievement Tests*. New York: College Entrance Examination Board.

Fleming, J. 1986. *Standardized Test Scores and the Black College Environment*. Research Report Volume 10 (Number 1). New York: United Negro College Fund.

Goldman, R., and B. N. Hewitt. 1976. Predicting the Success of Black, Chicano, Oriental, and White College Students. *Journal of Educational Measurement* 13 (2).

Kulik, J., R. Bangert-Drowns, and C. Kulik. 1984. Effectiveness of Coaching for Aptitude Tests. *Psychological Bulletin* 95:179–88.

Manski, C., and D. Wise. 1983. *College Choice in America*. Cambridge, Mass.: Harvard University Press.

Powers, D. E. 1988. *Preparing for the SAT: A Survey of Programs and Resources*. College Board Report 88-7. New York: College Entrance Examination Board.

Ramist, L., and S. Arbeiter. 1986. *Profiles, College-Bound Seniors, 1985*. New York: College Entrance Examination Board.

Valentine, J. A. 1987. *The College Board and the School Curriculum*. New York: College Entrance Examination Board.